

# Лекция 6: Методы численного моделирования биологических систем: Молекулярная Динамика и Динамика Ланжевена

## 1 Моделирование в полноатомном разрешении

Взаимоотношение между координатами атомов биомолекулы и её энергией является основной частью любых вычислительных исследований, основанных на атомарных моделях. Функция потенциальной энергии (силовое поле, force-field) [1, 2] используется для того, чтобы вычислить энергию системы, а также определить силы взаимодействия между входящими в неё атомами. Силовое поле задаётся функцией от координат атомов системы (функциональная часть силового поля), зависящей от некоторого набора параметров (параметрическая часть). Функциональная часть силового поля  $V$  обычно задаётся суммой ковалентных ( $V_b$ ) и нековалентных взаимодействий ( $V_{nb}$ ) в системе [3–7]:

$$V = V_b + V_{nb}, \quad (1)$$

где ковалентные взаимодействия описываются гармоническими вибрациями ковалентных связей (bonds), углов между тремя атомами (angles) и двумя разными типами торсионных взаимодействий - для правильных (proper dihedrals) и неправильных торсионных углов

(improper dihedrals):

$$V_b = \sum_{bonds} K_b(b-b_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} K_\phi(1-\cos(n\phi-\phi_0)) + \sum_{impropers} K_\psi(\psi-\psi_0)^2 \quad (2)$$

Например,  $b$  и  $\theta$  - расстояние между двумя атомами и угол между двумя смежными ковалентными связями,  $\phi$  и  $\psi$  - торсионные углы поворота двух связей вокруг третьей, смежной к ним.  $b_0$ ,  $\theta_0$ ,  $\phi_0$ , и  $\psi_0$  - равновесные значения этих величин. Эти значения рассчитываются из текущего расположения атомов в системе (их пространственных координат).  $K_b$ ,  $K_\theta$ ,  $K_\phi$  - коэффициенты упругости вибрации длинны ковалентной связи, угла между двумя смежными связями, правильных и неправильных торсионных вращений. Важно заметить, что данные параметры зависят от типов атомов и задаются изначально для каждого конкретного силового поля (параметрическая часть силового поля) [3, 8]. Это, теоретически, позволяет производить молекулярное моделирование для любой системы, вне зависимости от того, разрешена ли её кристаллическая структура. В некоторых силовых полях формула (2) дополняется дополнительными членами, например, корректировками Урея-Брэдли[3] и/или корректировками СМАР для торсионных углов [9, 10]. Нековалентная часть функции потенциальной энергии чаще всего состоит из потенциала электростатических взаимодействий и потенциала Ван-дер-Ваальса:

$$V_{nb} = \sum_{i,j} \left( \frac{q_i q_j}{4\pi\epsilon_0\epsilon r_{ij}} + \epsilon_{ij} \left[ \left( \frac{R_{ij}^{min}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}^{min}}{r_{ij}} \right)^6 \right] \right) \quad (3)$$

где  $r_{ij}$  - расстояние между двумя взаимодействующими атомами,  $q_i$  и  $q_j$  - заряды атомов;  $\epsilon$  и  $\epsilon_0$  - диэлектрическая проницаемость среды и диэлектрическая постоянная;  $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$  и  $R_{ij}^{min} = (R_i^{min} + R_j^{min})/2$  - параметры Ван-дер-Ваальса для атомов  $i$  и  $j$ . Изначально, функция потенциальной энергии также явно включала член, описывающие водородные связи [3], но позднее от него отказались, неявно включив данные взаимодействия в электростатический потенциал при помощи частичных зарядов на донорах и акцепторах водородных связей [4].

Существуют также модели, неявно описывающие взаимодействие с окружающей средой (вода) [11, 12]. Для этого потенциальная функция расширяется дополнительным членом, имитирующим гидрофобный эффект и заземление электростатики, а механические столкновения с молекулами воды моделируются при помощи случайной силы - уравнения Ньютона заменяются уравнениями Ланжевена. Самым популярным подходами для моделирования гидрофобного эффекта является энергетическая функция, пропорциональная площади атомов, доступной растворителю (модель SASA) [13–16]. Электростатические эффекты чаще всего описываются при помощи обобщённой теории Борна [16–20]. При моделировании в неявном растворителе общее количество степеней свободы уменьшается, в зависимости от размера системы, примерно в 10 раз, если сравнивать с моделированием в явном растворителе. Но, так как используется схожая форма потенциальной функции, шаг по времени по-прежнему не может превышать 1-2 фс.

## 2 Упрощённые модели биомолекул

Современные вычислительные возможности достаточны для достижения временных интервалов в сотни наносекунд даже при использовании полноатомного разрешения. Однако, большая часть процессов внутри клетки (таких как белок-белковые взаимодействия, смена конформаций белков) происходит на временных интервалах в микросекунды или даже миллисекунды. Разрешающая способность экспериментального оборудования, используемого для изучения единичных молекул (атомно-силовая микроскопия, эксперименты с оптическим пинцетом) также ограничивается миллисекундами-секундами. Поэтому, прямое сопоставление результатов моделирования в полноатомном разрешении с физиологическими процессами, а также их сравнение с доступными экспериментальными данными, не представляется возможным. С другой стороны, объёмные структуры большого числа молекулярных агрегатов и больших белковых образований были разрешены при помощи методов со слабым разрешением, таких, как CryoEM и рентгеновская дифракция на малых углах. Таким образом, идея использования упрощённого приближения путём объединения

большого числа степеней свободы в меньшее, возникает сама собой.

Упрощённые модели уже давно используются в молекулярном моделировании. На текущий момент, существует очень большое количество подходов как к упрощению моделей (уменьшению степеней свободы), так и к их параметризации. В некоторых случаях используется очень слабое упрощение, когда, например, на одну аминокислоту приходится 4-6 частиц, другие модели используют более грубое приближение, когда одна аминокислота описывается одной или двумя частицами. Есть подходы и с ещё более грубым упрощением, когда за структурную единицу берутся целые фрагменты белка. Как правило, более грубые приближения менее вычислительно затратны. Но, чем грубее модель, тем сложнее сделать её одновременно переносимой (применимой к множеству разных систем) и точной. Поэтому, более грубые модели зачастую отталкиваются от доступной нативной структуры изучаемого белка, используя её для параметризации.

### 3 Модель самоорганизующегося полимера (SOP)

Модель самоорганизующегося полимера (Self Organized Polymer, SOP) - упрощённая модель белков типа  $G\bar{b}$ , в которой каждая аминокислота представлена отдельным взаимодействующим центром [21]. Данная модель была разработана для описания механических свойств белков, моделируя процесс экспериментов на одиночных молекулах белка (таких как атомная силовая микроскопия и экспериментов при помощи оптических пинцетов). В более ранних исследованиях эта модель была успешно применена для описания зелёного флуоресцентного белка [22], кинезина [23] и димера тубулина [24]. За центры взаимодействия в модели SOP берутся атомы  $C_\alpha$ , а потенциальная энергия состояния белка зависит

от координат  $\{r\} = r_1, r_2, \dots, r_N$  этих атомов и задаётся выражением:

$$\begin{aligned}
V &= V_{FENE} + V_{NB}^{ATT} + V_{NB}^{REP} \\
&= - \sum_{\text{covalent}} \frac{k}{2} R_0^2 \log \left[ 1 - \frac{(r_{ij} - r_{ij}^0)^2}{R_0^2} \right] + \\
&+ \sum_{\text{native}} \varepsilon_h \left[ \left( \frac{r_{ij}^0}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{ij}^0}{r_{ij}} \right)^6 \right] + \\
&+ \sum_{\text{repulsive}} \varepsilon_r \left( \frac{\sigma}{r_{ij}} \right)^6
\end{aligned} \tag{4}$$

В уравнении 4, первое слагаемое описывает ковалентные связи в цепи при помощи конечно-растяжимого нелинейного эластичного потенциала (англ. *Finitely Extensible Nonlinear Elastic*, FENE)  $V_{FENE}$  [8]. Расстояние между соседними частицами  $i$  и  $j$  задаётся как  $r_{ij}$ , а  $r_{ij}^0$  - его равновесное значение (значение в нативном состоянии белка),  $R_0 = 2\text{\AA}$  - индекс чувствительности к изменениям ковалентной связи. Суммирование производится как по амидным связям белка, так и по дисульфидным связям. Второе слагаемое, описываемое потенциалом Леннарда-Джонса ( $V_{NB}^{ATT}$ ), используется для описания нековалентных связей, фиксирующих нативное состояние (нативных контактов). Частицы  $i$  и  $j$  образуют нативный контакт, если они не связаны ковалентными связями и при этом расположены достаточно близко (в пределах  $R = 8\text{\AA}$ ) друг от друга в нативной структуре. Значение параметра  $\varepsilon_h = 0,7 - 1,6$  ккал/моль характеризует силу нековалентных связей. Все прочие пары атомов (не связанные ковалентно и не формирующие нативный контакт) взаимодействуют согласно потенциалу  $V_{NB}^{REP}$ , благодаря которому становятся невозможными самопересечения цепи. Параметры  $\varepsilon_r = 1.0$  ккал/моль и  $\sigma = 3.8\text{\AA}$  постоянны и характеризуют силу и радиус отталкивания.

Основным преимуществом модели SOP является то, что не смотря на простоту силового поля, она достаточно точно описывает механическую денатурацию белков. Кроме того, благодаря использованию динамики Ланжевена в задемпфированном пределе, вычислительные процедуры остаются стабильными даже при использовании большого шага по времени в

40пс. Действительно, данная модель успешно была использована для достаточно больших систем без применения аппаратного ускорения [22, 24]. Схожая модель также применялась для симуляций капсида вируса ССМV на вычислительном кластере [25]. Однако, моделирование больших систем на экспериментальных временных интервалах с использованием центрального процессора невозможно даже с применением модели SOP (Рис. ??). Поэтому, в данной работе модель была реализована на графических процессорах, что позволило обойти данный барьер.

## Список литературы

- [1] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, “All-atom empirical potential for molecular modeling and dynamics studies of proteins,” *J. Phys. Chem. B*, vol. 102, no. 18, pp. 3586–3616, 1998.
- [2] A. D. MacKerell, “Empirical force fields for biological macromolecules: Overview and issues,” *J. Comput. Chem.*, vol. 25, no. 13, pp. 1584–1604, 2004.
- [3] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, “CHARMM: A programm for macromolecular energy, minimization, and dynamics calculations,” *J. Comput. Chem.*, vol. 4, no. 2, pp. 187–217, 1983.
- [4] B. R. Brooks, C. L. Brooks, III, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, “CHARMM: The biomolecular simulation program,” *J. Comput. Chem.*, vol. 30, no. 10, pp. 1545–1614, 2009.
- [5] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, “Scalable molecular dynamics with NAMD,” *J. Comput. Chem.*, vol. 26, pp. 1781–1802, 2005.
- [6] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, “GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation,” *J. Chem. Theory Comput.*, vol. 4, pp. 435–447, 2008.
- [7] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, “The Amber biomolecular simulation programs,” *J. Comput. Chem.*, vol. 26, no. 16, pp. 1668–1688, 2005.

- [8] D. van der Spoel, E. Lindahl, B. Hess, C. Kutzner, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen, and H. J. C. Berendsen, *GROMACS User Manual*. The GROMACS development team, 4.0 ed., 2009.
- [9] A. D. MacKerell, M. Feig, and C. L. Brooks, III, "Improved treatment of the protein backbone in empirical force fields," *JACS*, vol. 126, no. 3, pp. 698–699, 2004.
- [10] A. D. MacKerell, M. Feig, and C. L. Brooks, III, "Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations," *J. Comput. Chem.*, vol. 25, no. 11, pp. 1400–1415, 2004.
- [11] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules," *Nature Struct. Biol.*, vol. 9, no. 9, pp. 646–652, 2002.
- [12] J. Chen, C. L. Brooks, III, and J. Khandogin, "Recent advances in implicit solvent-based methods for biomolecular simulations," *Cur. Opin. Struct. Biol.*, vol. 18, no. 2, pp. 140–148, 2008.
- [13] P. Ferrara, J. Apostolakis, and A. Caflisch, "Evaluation of a fast implicit solvent model for molecular dynamics simulations," *Proteins*, vol. 46, no. 1, pp. 24–33, 2002.
- [14] F. Fraternali and W. F. van Gunsteren, "An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution," *J. Mol. Biol.*, vol. 256, no. 5, pp. 939–948, 1996.
- [15] D. Eisenberg and A. D. McLachlan, "Solvation energy in protein folding and binding," *Nature*, vol. 319, pp. 199–203, 1986.
- [16] D. Qiu, P. S. Shenkin, F. P. Hollinger, and W. C. Still, "The GB/SA continuum model for solvation. a fast analytical method for the calculation of approximate Born radii," *J. Phys. Chem. A*, vol. 101, no. 16, pp. 3005–3014, 1997.
- [17] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, "Semianalytical treatment of solvation for molecular mechanics and dynamics," *JACS*, vol. 112, no. 16, pp. 6127–6129, 1990.

- [18] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, "Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium," *J. Phys. Chem.*, vol. 100, no. 51, pp. 19824–19839, 1996.
- [19] B. N. Dominy and C. L. Brooks, III, "Development of a Generalized Born model parametrization for proteins and nucleic acids," *J. Phys. Chem. B*, vol. 103, no. 18, pp. 3765–3773, 1999.
- [20] A. Onufriev, D. Bashford, and D. A. Case, "Modification of the Generalized Born model suitable for macromolecules," *J. Phys. Chem. B*, vol. 104, no. 15, pp. 3712–3720, 2000.
- [21] C. Hyeon, R. I. Dima, and D. Thirumalai, "Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins," *Structure*, vol. 14, pp. 1633–1645, 2006.
- [22] R. I. D. M. Mickler, H. Dietz, C. Hyeon, D. Thirumalai, and M. Rief, "Revealing the bifurcation in the unfolding pathways of GFP using single molecule experiments and simulations," *Proc. Natl. Acad. Sci. USA*, vol. 104, pp. 20268–20273, 2007.
- [23] T. Veitshans, D. Klimov, and D. Thirumalai, "Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties," *Fold. Des.*, vol. 2, no. 1, pp. 1–22, 1997.
- [24] R. I. Dima and H. Joshi, "Probing the origin of tubulin rigidity with molecular simulations," *Proc. Natl. Acad. Sci. USA*, vol. 105, no. 41, pp. 15743–15748, 2008.
- [25] M. Cieplak and M. O. Robbins, "Nanoindentation of virus capsids in a molecular model," *J. Chem. Phys.*, vol. 132, no. 1, p. 015101, 2010.